REMARKS

Claims 1, 3-6, 10-19, and 28-39 were pending in the present application. Independent Claims 1, 28, and 36 have been amended to clarify claimed subject matter and/or correct informalities. Support for these amendments may be found in the original Specification at least on page 15, lines 4-12; page 17, lines 10-20; page 18, lines 17-25; page 19, lines 19-22; and Figures 2, 3, 4, and 5.

Claims 40-44 are newly added. Support for the claim additions can be found in the original Specification at least at page 19, lines 10-12 and Figure 4. Claims 40-44 provide additional scope of subject matter commensurate with the original disclosure.

Claims 1, 3-6, 10-19, and 28-44 are for consideration upon entry of the present amendment. No new matter has been introduced by these amendments. Applicant requests favorable consideration of this response and allowance of the subject application based on the following remarks.

Statement of Substance of Interview

Applicant appreciates the Office's participation in a telephonic conference of July 25, 2006.

During the interview, the claimed subject matter of the application and the Ramaswamy reference were discussed. Specifically, Applicant presented arguments as to how Ramaswamy lacked features of the claimed subject matter, such as training unit, tuning set, and empirically derived value.

Also discussed during the interview were proposed amendments to the claims. In the interest of expediting prosecution of the application, and without conceding the

propriety of the rejection, Applicant proposed to amend each of the independent claims to further clarify features of the claimed subject matter. The Examiner requested that the proposed amendments be presented in writing. Accordingly, Applicant is submitting the proposed amendments in writing in this response to the Office Action.

<u>**Claim Rejections 35 U.S.C. §102**</u>

Claims 1, 3-6, 14-19, 28-30, 32, 33, and 35-39

Claims 1, 3-6, 14-19, 28-30, 32, 33, and 35-39 stand rejected under 35 U.S.C. §102(e) as being allegedly anticipated by U.S. Patent No. 6,188,976 to Ramaswamy et al. (hereinafter "Ramaswamy"). Applicant respectfully traverses this rejection. Anticipation under §102 requires that each and every element as set forth in the claim is found, either expressly or inherently described, in a single prior art reference (MPEP §2131).

Independent Claim 1 has been amended to further clarify features of Applicant's subject matter. Claim 1 now recites:

> *A method of using a tuning set of information to jointly optimize the performance and size of a language model, comprising:*
> *providing a textual corpus comprising subsets wherein each subset comprises a plurality of items;*
> *creating a Dynamic Order Markov Model data structure by assigning each item of the plurality of items to a node in the data structure, wherein the nodes are logically coupled to denote dependencies of the items, and calculating a frequency of occurrence for each item of the plurality of items;*
> *segmenting at least a subset of a received textual corpus into segments by clustering every N-items of the received corpus into a training unit, wherein resultant training units are separated by gaps, and wherein N is an empirically derived value based, at least in part, on the size of the received corpus;*
> *creating the tuning set from application-specific information;*
> *(a) training a seed model via the tuning set;*
> *(b) calculating a similarity within a sequence of the training units on either side of each of the gaps;*
> *(c) selecting segment boundaries that maximize intra-segment similarity and inter-segment disparity;*

(d)calculating a perplexity value for each segment based on a comparison with the seed model;
(e)selecting some of the segments based on their respective perplexity values to augment the tuning set;
iteratively refining the tuning set and the seed model by repeating parts (a) through (e) with respect to a threshold; and
refining the language model based on the seed model.

Ramaswamy does not disclose expressly or inherently "creating a Dynamic Order Markov Model data structure by assigning each item of the plurality of items to a node in the data structure, wherein the nodes are logically coupled to denote dependencies of the items and calculating a frequency of occurrence value for each item of the plurality of items", as recited in Claim 1. While Ramaswamy describes a domain-specific language model based on a relatively small seed group of linguistic units relevant to the domain (col. 2, lines 4-6) and extraction of linguistic units (col. 3, lines 42-44), this evidence is insufficient to support a prima facie anticipation rejection of the claimed Dynamic Order Markov Model data structure where each item of a subset is assigned to a node.

Consequently, Applicant respectfully submits that Claim 1 is not anticipated by Ramaswamy and requests that the §102 rejection be withdrawn.

Dependent Claims 3-6 and 10-19 depend directly or indirectly from Claim 1 and thus are allowable as depending from an allowable base claim. These claims are also allowable for their own recited features that, in combination with those recited in Claim 1, are not disclosed by Ramaswamy.

Independent Claim 28 recites:

A modeling agent comprising:
a controller, to receive invocation requests to develop a language model from a textual corpus comprising subsets wherein each subset comprises a plurality of items, and to calculate a frequency of occurrence for each item of the plurality of items;
a data structure generator, responsive to the controller, to:

create a Dynamic Order Markov Model data structure by assigning
each item of the plurality of items to a node in the data structure, wherein
the nodes are logically coupled to denote dependencies of the items;
develop a seed model from a tuning set of information;
segment at least a subset of a received corpus, wherein the segments
of the received corpus are a clustering of every N items of the received
corpus into a training unit, wherein N is an empirically derived value
based, at least in part, on the size of the received corpus, and the training
units are separated by gaps;
calculate the similarity within a sequence of training units on either
side of each of the gaps;
select segment boundaries that improve intra-segment similarity and
inter-segment disparity;
calculate a perplexity value for each segment;
refine the seed model with one or more segments of the received
corpus based, at least in part, on the calculated perplexity values;
iteratively refine the tuning set with segments ranked by the seed
model and in turn iteratively update the seed model via the refined tuning
set;
filter the received corpus via the seed model to find low-perplexity
segments; and
train the language model via the low-perplexity segments.

Independent Claim 28 is amended to recite features similar to those in Claim 1 and

hence benefits from the same arguments directed above to Claim 1.

In addition, Ramaswamy does not disclose expressly or inherently "a controller, a

data structure generator, or a Dynamic Order Markov Model data structure", as recited in

Claim 28.

Applicant asserts Ramaswamy fails to anticipate independent Claims 1 and 28

because Ramaswamy does not disclose the recited features of the claimed subject matter.

Accordingly, Applicant requests that the §102 rejections be withdrawn.

Dependent Claims 29-35 depend directly or indirectly from Claim 28 and thus are

allowable as depending from an allowable base claim. These claims are also allowable for

their own recited features that, in combination with those recited in Claim 28, are not disclosed by Ramaswamy.

Independent Claim 36 recites:

A method of jointly optimizing the performance and size of a language model, comprising:

providing a textual corpus comprising subsets wherein each subset comprises a plurality of items;

creating a Dynamic Order Markov Model data structure by assigning each item of the plurality of items to a node in the data structure, wherein the nodes are logically coupled to denote dependencies of the items, and calculating a frequency of occurrence for each item of the plurality of items;

segmenting one or more relatively large language corpora into multiple segments of N items, wherein N is an empirically derived value based, at least in part, on the size of the received corpus;

selecting an initial tuning sample of application-specific data, the initial tuning sample being relatively small in comparison to the one or more relatively large language corpora, wherein the initial tuning sample is used for training a seed model, the seed model to be used for ranking the multiple segments from the language corpora;

iteratively training the seed model to obtain a mature seed model, wherein the iterative training proceeds until a threshold is reached, each iteration of the training including:

updating the seed model according to the tuning sample;

ranking each of the multiple segments according to a perplexity comparison with the seed model;

selecting some of the multiple segments that possess a low perplexity; and

augmenting the tuning sample with the selected segments;

once the threshold is reached, filtering the language corpora according to the mature seed model to select low-perplexity segments;

combining data from the low-perplexity segments; and

training the language model according to the combined data.

Independent Claim 36 is amended to recite features similar to those in Claim 1 and hence benefits from the same arguments directed above to Claim 1. Applicant asserts Ramaswamy fails to anticipate independent Claims 1 and 36 because Ramaswamy does

not disclose the recited features of the claimed subject matter. Accordingly, Applicant requests withdrawal of the §102 rejections of these claims.

Dependent Claims 37-39 depend from Claim 36 and are allowable as depending from an allowable base claim. These claims are also allowable for their own recited features, which, in combination with those recited in Claim 36, are neither shown nor suggested by Ramaswamy. Accordingly, Applicant requests withdrawal of the §102 rejections of these claims.

Claim Rejections 35 U.S.C. §103

Claims 10-13, 31 and 34

Claims 10-13, 31 and 34 stand rejected under 35 U.S.C. §103(a) as being unpatentable under Ramaswamy in view of U.S. Patent No. 6,317,707 to Bangalore et al. (hereinafter "Bangalore"). Applicant respectfully traverses the rejections. To establish a prima facie case of obviousness, three basic criteria must be met. First, there must be some suggestion or motivation, either in the references themselves or in the knowledge generally available to one of ordinary skill in the art, to modify the reference or to combine reference teachings. Second, there must be a reasonable expectation of success. Finally, the prior art reference (or references when combined) must teach or suggest all the claim limitations (see, MPEP 2142).

Dependent Claim 10 recites in part, *"the calculation of the similarity within a sequence of training units defines a cohesion score."*

Dependent Claim 11 recites in part, *"intra-segment similarity is measured by cohesion score."*

Dependent Claim 12 recites in part, *"inter-segment disparity is approximated from the cohesion score."*

Dependent Claim 13 recites in part, *"the calculation of inter-segment disparity defines a depth score."*

Dependent Claims 10-13 depend directly or indirectly from Claim 1 and hence benefit from the same arguments directed above to Claim 1. As explained above with respect to the rejection under 35 U.S.C. § 102(e), Applicant submits that Ramaswamy does not disclose "creating a Dynamic Order Markov Model data structure by assigning each item of the plurality of items to a node in the data structure, wherein the nodes are logically coupled to denote dependencies of the items and calculating a frequency of occurrence for each item of the plurality of items", as recited in Claim 1.

The Office states Ramaswamy does not mention the features of Claim 10 but that Bangalore teaches this subject matter (Office Action, page 10). Bangalore does not compensate for the deficiencies of Ramaswamy, as neither reference discloses, teaches, or suggests the recited features of Claims 1 and 10. Instead, Bangalore describes a "close relationship" (col. 3, lines 15-19), not a cohesion score. Bangalore merely shows how input words having the same lexical significance should possess similar vectors in the frequency space (col. 3, lines 3-4), not the calculation of similarity within a sequence of training units. Thus, Bangalore does not disclose, teach, or suggest this feature.

The Office has failed to establish a motivation sufficient for one of ordinary skill in the art to combine these references. The motivation provided by the Office to "determine how close or similar the training units were to each other for the benefit of maximizing the clustering process of related items" is very general because it could cover almost any

alteration contemplated of Ramaswamy. Additionally, there is nothing in either references that would suggest a 'cohesion score' to calculate the similarity within a sequence of training units. Finally, although Bangalore describes a close relationship, there is no suggestion, other than Applicant's disclosure, to employ this close relationship to calculate the similarity within a sequence of training units. This rejection is improper.

Since these features are not disclosed, taught, or suggested in Ramaswamy or Bangalore alone, the resultant combination of these references does not result in Applicant's Claims 10, 11, 12, and 13. Accordingly, these claims are allowable over these references, individually or in combination for at least these reasons.

Dependent Claims 11 and 12 depend directly or indirectly on Claims 1 and 10, and **Dependent Claim 13** depends directly on Claim 1, and these claims are believed patentable for at least the same reasons. Applicant respectfully requests that the rejection of these claims be withdrawn.

Dependent Claims 31 and 34 recite in part, "*a frequency analysis function, to determine a frequency of occurrence of segments within the received corpus*".

Dependent Claims 31 and 34 depend directly or indirectly from Claim 28 and hence benefit from the same arguments directed above to Claim 28. As explained above with respect to the rejection under 35 U.S.C. § 102(e), Applicant submits that Ramaswamy does not disclose "a controller, an analysis engine, a data structure generator, or a Dynamic Order Markov Model data structure, wherein the nodes are logically coupled to denote dependencies of the items", as recited in Claim 28.

The Office states Ramaswamy does not mention the features of Claim 31 but that Bangalore teaches this subject matter (Office Action, page 13).

Bangalore does not compensate for the deficiencies of Ramaswamy, as neither reference teaches or suggests the recited features of Claims 28, 31, and 34. Instead, Bangalore describes a single, all encompassing cluster enclosing all clusters and input words (col. 3, lines 45-48). Bangalore shows forming a root cluster (col. 3, line 52), not determining a frequency of occurrence of segments within the received corpus. Thus, Bangalore does not disclose, teach, or suggest this feature.

The Office has failed to establish a motivation sufficient for one of ordinary skill in the art to combine these references. The motivation provided by the Office "for the benefit of maximizing the clustering segments to better improve subsequent language modeling results", is too general because it could cover almost any alteration contemplated of Ramaswamy. Additionally, there is nothing in either references that would suggest 'maximizing clustering segments' would improve language modeling results. Furthermore, there is no suggestion by the references, other than Applicant's disclosure, to employ a frequency analysis function to determine a frequency of occurrence of segments within the received corpus. This rejection is improper. The Office cannot improperly rely on hindsight without evidence of motivation to propose the suggested combination. Applicant respectfully requests the §103 rejection be withdrawn.

Since these features are not disclosed, taught, or suggested in Ramaswamy or Bangalore alone, the resultant combination of these references does not result in Applicant's Claims 31 and 34. Accordingly, these claims are allowable over these references, individually or in combination for at least these reasons. Applicant respectfully requests that the rejection of these claims be withdrawn.

New Claims

Claims 40-44 are newly added. Support for the claim additions can be found in the original Specification at least at page 19, lines 10-12 and Figure 4. Claims 40-44 provide additional scope of subject matter commensurate with the original disclosure.

Conclusion

All pending and new claims are in condition for allowance. Applicant respectfully requests reconsideration and prompt issuance of the present application. If any issues remain that preclude issuance of the application, the Examiner is urged to contact the undersigned attorney before issuing a subsequent Action.

Respectfully Submitted,

Dated: 8-15-2006   By: *Shirley Lee Anderson*
Shirley Lee Anderson
Reg. No. 57,763
(509) 324-9256 ext. 258
Lee & Hayes, PLLC